Remarks

The following remarks are to address rejections made in Application Serial No. 09/696,953 from which the present application is a continuation application.

Claims 1, 2, 4, 7, 10, 11, 13, 16, 19, 20, 22 and 25 of the parent application were rejected under 35 U.S.C. §103(a) as being unpatentable over Baker (US 4,803,729) in view of Marley (US 4,783,807) and DiRonza (US 5,644,678). Claims 3, 12 and 21 of the parent application were rejected under 35 U.S.C. §103(a) as being unpatentable over Baker in view of Marley and DiRonza and further in view of Moshier (US 4,489,434).

It is apparent that the claims filed with this application are patentable over these rejections for the following reasons.

Claim 1 is patentable over the combination of Baker, Marley and DiRonza, since claim 1 recites a voice pitch normalization apparatus having, in part, a voice pitch normalization device operable to generate a target voice signal by changing an incoming command voice signal on a predetermined degree basis and operable to change the target voice signal in voice pitch unit until a maximum of probabilities indicating a degree of coincidence among the target voice signal and a plurality of words in sample data reaches a predetermined probability or higher. The combination of Baker, Marley and DiRonza fails to disclose or suggest a voice pitch normalization apparatus as recited in claim 1.

Baker discloses a speech recognition method 40. In the method 40, the first step 42 converts speech to be recognized into a sequence of acoustic frames. The acoustic frames give the acoustic parameters of the speech during a short period of time. A second step 44 associates a phonetic label with each frame of the speech to be recognized. This association is referred to as 'smooth frame labeling.' A third step 46 divides the speech into segments of frames which have phonetic labels of the same class (i.e., into segments in which the frames are all either associated with consonant sounds, vowel sounds, or silence sounds). A fourth step 48 determines which of a number of previously calculated diphone-type models match best against the frames on each side of the boundaries between the segments detected in the third step 46. Each diphone-type model contains two sub-models, one representing sound before and the other representing sound after.

In the method 40, a fifth step 50 produces for each of the diphone-type models associated with each segment boundary, a displaced evidence score for each word in which the diphone-type model is known to occur. A sixth step 52 derives a histogram score for each vocabulary word in association with each of a plurality of blocks of frames of speech. Once the evidence scores and the histogram scores have been calculated for a block of frames, a seventh step 54 combines, for each word, all the displaced evidence scores for that word and the histogram score for that word associated with that block. An eighth step 56 then selects for each block of frames, a number of word candidates that have a probability of corresponding to that block of frames. Finally, a ninth step 58 utilizes dynamic programming against the word candidates to determine which of the word candidates has the highest probability of corresponding to the actual speech. (See column 11, line 39 - column 12, line 23 and Figure 1).

Based on the above discussion, it is clear the speech recognition method of Baker operates by segmenting the speech inputted thereto into a number of frames and the frames are compared to a number of models on a frame-by-frame basis and also compared to a words on a block of frames-by-block of frames basis. The method then determines which word has the highest probability of corresponding to the actual speech based on scores determined by these two types of comparisons and selects this word. The word with the highest probability does not have to meet any threshold value. Baker does not disclose or suggest generating a target voice signal by changing the incoming voice signal on a predetermined degree basis, or changing the target voice signal in voice pitch unit until a maximum of probabilities indicating a degree of coincidence among the target voice signal and a plurality of words in sample data reaches a predetermined probability or higher.

Marley discloses a system for sound recognition with feature selection synchronized to voice pitch. The system operates by producing a binary signal having a "1" level during positive pressure wave portions of a sound signal and a "0" level during negative pressure wave portions of the sound signal, detecting a time point of major peak positive and negative excursions of each pitch cycle (i.e., the period of time of adjacent positive and negative pressure wave portions) of the sound signal and producing a corresponding pitch cycle marker signal that occurs at the beginning of each pitch cycle, producing a first number that represents the duration of a "1" level of the binary signal most closely following a pulse of the pitch cycle signal and producing a second number that represents the duration

of the following "1" level of the binary signal, composing a vector from the first and second numbers, comparing the vector with a plurality of stored vector domains to determine if the present vector falls within any of the stored vector domains, and producing a character signal representing a phoneme or sound corresponding to one of the stored vectors that most nearly matches the present input vector. Further, vectors based on running averages of the durations of "1"s and "0"s are calculated to compare with stored vectors. (See column 4, lines 1-22).

As can be seen from the above discussion, Marley relies on the values of vectors that are created by calculating the difference between durations of two adjacent positive pressure wave portions of a sound signal and running averages of the durations of positive and negative pressure waves to match the sound signal with previously stored vectors corresponding to a phoneme or sound. While Marley does recognize that the normalization of speech signals to compensate for amplitude and pitch variations is a problem that needs to be addressed (see column 1, lines 36-48), Marley fails to disclose or suggest a voice pitch normalization device operable to generate a target voice signal by changing an incoming command voice signal on a predetermined degree basis and operable to change the target voice signal in voice pitch unit until a maximum of probabilities indicating a degree of coincidence among the target voice signal and a plurality of words in sample data reaches a predetermined probability or higher. Instead, Marley relies on the generation of vectors as described above to match a sound signal with previously stored information and not on changing the pitch of the sound signal on a predetermined basis.

DiRonza discloses a method of estimating a pitch of a speech acoustic signal in a time interval. (See column 1, line 43 - column 2, line 22). However, while DiRonza discloses a method to calculate the pitch of a signal, it is apparent that it fails to disclose or suggest a voice pitch normalization device operable to generate a target voice signal by changing an incoming command voice signal on a predetermined degree basis and operable to change the target voice signal in voice pitch unit until a maximum of probabilities indicating a degree of coincidence among the target voice signal and a plurality of words in sample data reaches a predetermined probability or higher.

Based on the above discussion of the references relied upon in the rejection, it is apparent that the combination of these references fails to disclose or suggest a voice pitch normalization device as recited in claim 1.

In section 9 of the Office Action, Moshier is relied upon as disclosing a timing clock. However, even if this contention is accurate, Moshier also fails to disclose or suggest above-discussed features of claim 1.

As for claims 6 and 11, they are patentable over the references for similar reasons as set forth above with regard to claim 1. That is, claims 6 and 11, like claim 1, recite the generation of a target voice signal by changing an incoming command voice signal on a predetermined degree basis and the changing of the target voice signal in voice pitch unit until a maximum of probabilities indicating a degree of coincidence among the target voice signal and a plurality of words in sample data reaches a predetermined probability or higher, which features are not disclosed or suggested in any of the references, either individually or in combination.

Because of the above mentioned distinctions, it is believed clear that claims 1-15 are allowable over the references relied upon in the rejections. Furthermore, it is submitted that the distinctions are such that a person having ordinary skill in the art at the time of invention would not have been motivated to make any combination of the references of record in such a manner as to result in, or otherwise render obvious, the present invention as recited in claims 1-15. Therefore, it is submitted that claims 1-15 are clearly allowable over the prior art of record.

In view of the above remarks, it is submitted that the present application is now in condition for allowance. The Examiner is invited to contact the undersigned by telephone if it is felt that there are issues remaining which must be resolved before allowance of the application.

Respectfully submitted,

Mikio ODA et al.

By David M. Ovedovitz

David M. Ovedovitz
Registration No. 45,336
Attorney for Applicants

DMO/jmj
Washington, D.C. 20006-1021
Telephone (202) 721-8200
Facsimile (202) 721-8250
December 3, 2003